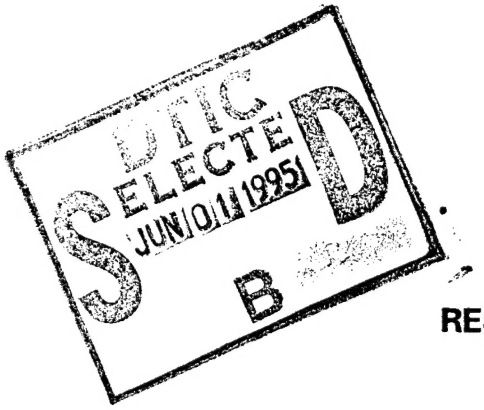
EDGEWOOD
RESEARCH, DEVELOPMENT & ENGINEERING CENTER
U.S. ARMY CHEMICAL AND BIOLOGICAL DEFENSE COMMAND

ERDEC-TR-227

HYPERACTIVATED RABBIT SPERM CELL MOTILITY PARAMETERS



B.A. Bodt
R.J. Young

RESEARCH AND TECHNOLOGY DIRECTORATE

March 1995

Aberdeen Proving Ground, MD 21010-5423

Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1995 March	Final, 91 Jun - 92 Sep

4. TITLE AND SUBTITLE

Hyperactivated Rabbit Sperm Cell Motility Parameters

5. FUNDING NUMBERS

PR-1N6A

6. AUTHOR(S)

Bodt, B.A., and Young, R.J.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

DIR, ERDEC,* ATTN: SCBRD-RTL, APG, MD 21010-5423

8. PERFORMING ORGANIZATION REPORT NUMBER

ERDEC-TR-227

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

*When this study was conducted, ERDEC was known as the U.S. Army Chemical Research, Development and Engineering Center, and the authors were assigned to the Research Directorate.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A variety of statistical procedures was used to analyze the motility parameters of separate populations of hyperactivated and non-hyperactivated rabbit sperm cells. The parameter Wobble (WOB) was the most efficient in classifying hyperactivation. In combination with Curvilinear Velocity (VC), at least 98% of the cells were correctly classified as either hyperactivated or non-hyperactivated. The threshold values for the two motility parameters were specific for the instrument used to measure the motion parameter. The ability to objectively identify and quantify hyperactivated motility is potentially of great use for clinical and toxicological assessment of fertility.

14. SUBJECT TERMS

Rabbit	Motility	Toxicity tests
Sperm cells	Chemical	Videotape
Hyperactivated	Toxicant	

15. NUMBER OF PAGES 37

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Blank

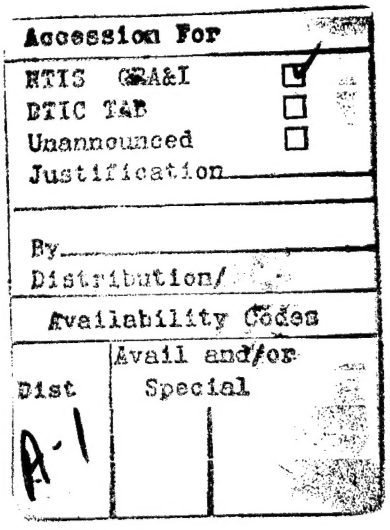
PREFACE

The work described in this report was authorized under Project No. 1N6A. This work was started in June 1991 and completed in September 1992.

In conducting the research described in this report, the investigators adhered to the "Guide for the Care and Use of Laboratory Animals," National Institute of Health Publication No. 86-23, 1985, as promulgated by the committee on Revision of the Guide for Laboratory Animal Facilities and Care of the Institute of Laboratory Animal Resources, Commission of Life Sciences, National Research Council (Washington, DC). These investigations were also performed in accordance with the requirements of AR 70-18, Laboratory Animals, Procurement, Transportation, Use, Care, and Public Affairs.

The use of trade names or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

This report has been approved for release to the public. Registered users should request additional copies from the Defense Technical Information Center; unregistered users should direct such requests to the National Technical Information Service.



QUALITY ASSURANCE

This study, governed by Protocol Number 210910430000, was examined for compliance with Good Laboratory Practices as published by the U.S. Environmental Protection Agency in 40 CFR Part 792 (effective 18 September 1989). The dates of all inspections and the dates the results of those inspections were reported to the Study Director and management were as follows:

Phase Inspected	Date Inspected	Date Reported to Study Director/Management
Videotaping	05 Nov 1991	05 Nov 1991
Final Report	18 July 1994	15 Sept 1994

To the best of my knowledge, the methods described in this report were the methods followed during the study as indicated by the raw data found in the laboratory notebook. The report was determined to be an accurate reflection of the raw data recorded.

Kenneth P. Cameron
Quality Specialist
Life Sciences Department

15 Sept 1994
Date

CONTENTS

FIGURES

TABLES

HYPERACTIVATED RABBIT SPERM CELL MOTILITY PARAMETERS

1. INTRODUCTION

A recent study showed that hyperactivated motility of rabbit sperm cells was suppressed by metals implicated in fertility disturbances and not by others devoid of this property.[1] Estimations of hyperactivated motility decrease were based on subjective visual observations, and although the decrease was found in several replicates, an objective method to measure hyperactivated motility is desirable. Hyperactivated motility is necessary for fertilization. To understand the mechanism underlying this phenomenon,[2] to apply it to assessing fertility effects associated with sperm cells' exposure to chemicals,[1] or to use it for fertility prediction in clinical settings[3,4] by measuring the decline in hyperactivation, great care must be taken in developing objective, accurate, and dependable rules for classifying hyperactivated and non-hyperactivated sperm. Statistical, analytical methods, based on the motion parameters determined by motion analytical systems, were brought to bear on the problem of identifying either the hyperactivated or non-hyperactivated state of individual sperm cells in a mixed population. Detailed statistical analyses were used to investigate and understand the relationship between the components of flagellar motion, their interrelationship, and their relationship to hyperactivity. Cell state was modeled as a function of motion parameter values, and model effectiveness was assessed in terms of misclassification error. The results of the investigation are presented in this report.

2. MATERIALS AND METHODS

2.1 Analysis of Sperm Cells.

Videotapes of the motion of rabbit sperm cells that did not develop hyperactivated motility after incubation for 1 or 2 hr, and those that developed hyperactivated motility after 16-20 hr incubations[5,6] were used for analyses. Analysis with the CellSoft system and methods for developing hyperactivated motility were carried out as previously described.[5-7] The settings for the CellTrak system (Motion Analysis Corporation, Santa Rosa, CA) were frame rate, 30 frames/s; duration of frame capture, 30 frames; minimum path length, 15 frames; minimum burst speed, 20 μm/s; maximum burst speed, 500 μm/s; distance scale factor, 1.839 μm/pixel; camera aspect ratio, 1.0; amplitude of lateral head (ALH) path smoothing factor, 7 frames; centroid X and Y search neighborhood, 4 and 2 pixels, respectively; centroid cell size minimum and maximum, 2 and 25 pixels, respectively; maximum path interpolation, 1 frame; path prediction percentage, 0%.

Hyperactivated sperm cells were identified using criteria previously defined.[8,9] When necessary, close visual inspection of the videotape was carried out frame by frame to ensure correct classification of the motility type.

The motion parameters measured were curvilinear velocity (VC), straight line velocity (VST), linearity (LIN), maximum amplitude of lateral head (MALH) displacement, average amplitude of lateral head (AALH) displacement, beat cross frequency (BCF), straightness (STR), wobble (WOB), AALH/LIN, and VC x AALH.

2.2 Statistical Procedures.

The statistical analysis was completed in the following four stages:

- Univariate examination of each motility characteristic between the classes of hyperactivated and non-hyperactivated motility was based on sample means, standard deviations, relative frequency distributions, and boxplots. An indication of variable importance was obtained by using the p-values for the Mann-Whitney test.

- Joint contribution of variables to classification were explored graphically using scatter plots provided by NCSS version 5.1, 1987, and BMDP 1983, program 6D.

- Classification was pursued using standard discriminant analysis and newer tree structured methods with available software. Stepwise discriminant analysis, complimented by binary regression, was performed using BMDP statistical software (BMDP 1983, programs 7M, 1R, and 9R).

- The Classification And Regression Trees (CART™, Version 1.1, California Statistical Software, Inc., Belmont, CA) and A Fast Algorithm for Classification Trees (FACT, Version 1.1, Software Development and Distribution Center, MACC, University of Wisconsin, Madison, WI) software were used to establish a decision tree for classification. CART™ was principally used with the FACT results serving to corroborate. Final results for misclassification errors were computed using cross validation.

3. RESULTS AND DISCUSSION

3.1 Motion Parameter Statistics.

Summary statistics for each of the 10 motion parameters for 322 hyperactivated and 899 non-hyperactivated sperm cells are given in Table 1. More detailed information is given in histograms appearing in the Appendix. The sample mean and standard deviation give an indication of where the center portion of the data lies, and the extreme points bound the values observed. Some unusual values are found in the table. The maximum for VC, MALH, and AALH is more than 5 standard deviations from the mean, and for AALH/LIN and VC*AALH, the maximum is more than 18 and 9 standard deviations, respectively. Hyperactivated cells generally show smaller values for VST, LIN, BCF, STR, and WOB (Table 1). For all motion parameters but LIN, the standard deviation differs between classes; in particular, note MALH, AALH, AALH/LIN, and VC*AALH.

Table 1. Summary Statistics for the Motility Parameters Determined for both Hyperactivated and Non-hyperactivated Cells

Motility Parameter	Class	Sample Mean ± SD	Minimum	Maximum	Range
VC	hyper	137.6 ± 52.0	51.0	344.8	293.8
	non-hyper	83.1 ± 35.7	23.2	191.3	168.1
VST	hyper	30.4 ± 21.1	0.1	104.4	104.3
	non-hyper	71.0 ± 35.5	0.7	175.5	174.8
LIN	hyper	0.24 ± 0.18	0.01	0.74	0.73
	non-hyper	0.85 ± 0.18	0.02	0.99	0.97
MALH	hyper	9.9 ± 4.9	0.2	27.8	27.6
	non-hyper	2.3 ± 1.3	0.6	10.9	10.3
AALH	hyper	7.1 ± 3.3	1.1	20.1	19.0
	non-hyper	1.6 ± 0.9	0.4	10.9	10.5
BCF	hyper	12.2 ± 5.5	1.2	25.9	24.7
	non-hyper	15.0 ± 3.8	1.2	27.2	26.0
STR	hyper	0.58 ± 0.30	0.02	0.98	0.96
	non-hyper	0.90 ± 0.14	0.04	0.99	0.95
WOB	hyper	0.40 ± 0.15	0.01	0.79	0.78
	non-hyper	0.94 ± 0.08	0.23	1.00	0.77
AALH/LIN	hyper	92.5 ± 192.2	2.2	2008.0	2005.8
	non-hyper	2.4 ± 3.7	0.4	53.5	53.1
VC*AALH	hyper	1124.3 ± 984.3	96.4	6459.7	6363.3
	non-hyper	144.9 ± 128.3	16.4	722.9	706.5

Summary statistics for each motility parameter are reported individually for each cell state. The mean ± the sample standard deviation, for the population, gives information as to the location of the majority of motility parameter values. The minimum, maximum, and range provide information as to the extremes. All summary statistics were computed using BMDP statistical software.

Data in Table 1 suggest that the differences in means and standard deviations would be statistically significant. This was confirmed by the nonparametric Mann-Whitney Test for location and the Squared Ranks test for variances ($p < 0.01$).[10] The caveat to this is that the enormous sample sizes (322 hyperactivated cells and 899 non-hyperactivated cells) will cause the power of the test to be quite high, even for relatively small differences between the hypothesized value of the parameter in question and its alternative.

3.2 Single Classification.

A difference in distribution location between motility classes for a motility parameter only hints that the parameter might be useful in classification. The extent to which the distributions overlap must be examined, because it is within the intervals where overlap occurs that the potential exists for misclassification. Figure 1 illustrates the overlap for each of the motility parameters using stacked boxplots of the hyperactivated (H) and non-hyperactivated (N) class distributions. The basic form of the boxplot, consisting of the quartiles and the minimum and maximum values, was used. For this figure, all motility parameter values were standardized, using the combined data mean and standard deviation for the scaling. This permitted the simultaneous viewing of the distributions for each parameter and comparison of all regarding their potential for use as classifiers. The numerical value is given for standardized values beyond four standard deviations from the mean. Figure 1 shows that for the parameters LIN and WOB, at most, 25% of the non-hyperactivated cells show values that are similar to those of the hyperactivated class. It is likely that AALH, MALH, and VC*AALH will also be reasonable classifiers, based on the degree of separation of motility classes seen between the boxes representing the middle 50% of the data. The BCF provides an example of a parameter with limited classifying potential.

The relative frequency distributions for LIN, VC, AALH, and WOB are compared between the two motility classes in Tables 2 and 3. Linearity, VC, and AALH were selected because of their prominence in the literature,[2-4] and WOB was selected for its importance in this study. Hyperactivated cells were absent in the 0.8 - 1.0 interval for both LIN and WOB (Table 2), and conversely high percentages of non-hyperactivated cells, LIN, 75.5% and WOB, 94.3% were found within this interval. This strongly suggests good classifying potential for each. The AALH shows only minimal distribution overlap, and VC has somewhat more. The individual concomitants of hyperactivation suggested by Mortimer and Mortimer for human sperm[11] are consistent with these results despite the fact that rabbit sperm values are reported here.

3.3 Multiple Classification.

The scatterplots in Figure 2 show the relationship between the paired values of the four motility parameters discussed above and each motility class. Each possible pairing for VC, LIN, and AALH is represented, as well as the pairing for VC and WOB. The symbol, h, indicates the presence of one or more hyperactivated cells with values of the two motility parameters defining its position; c denotes non-hyperactive cells, and an asterisk

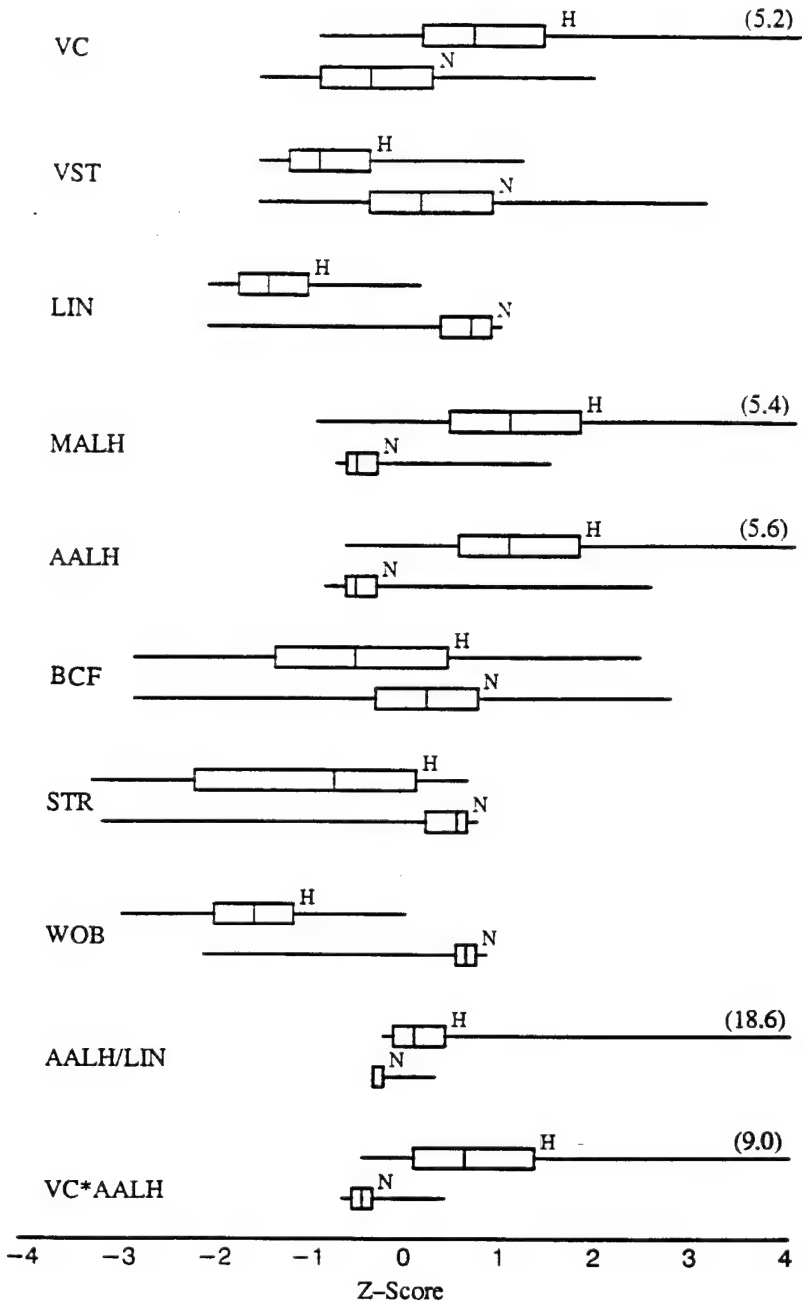


Figure 1. Boxplots of the Standardized Motility Parameter Distributions. The graphical summary allows a quick comparison of all motility parameters in their ability to separate on the basis of hyperactivation. The box is formed from the first and third quartiles, with the median indicated as a vertical line within the box. The extremes are connected to the box with a line segment.

Table 2. Relative Frequency Distributions (Given in Percents) of WOB and LIN for both Hyperactivated (n=322) and Non-hyperactivated (n=899) Cells

	LIN			WOB	
Interval	Hyper	Non-hyper	Interval	Hyper	Non-hyper
0.0 - 0.1	25.8	0.8	0.0 - 0.1	1.2	0.0
0.1 - 0.2	24.2	1.1	0.1 - 0.2	5.9	0.0
0.2 - 0.3	19.3	0.8	0.2 - 0.3	23.0	0.4
0.3 - 0.4	12.4	1.1	0.3 - 0.4	27.7	0.0
0.4 - 0.5	9.3	2.2	0.4 - 0.5	18.3	0.3
0.5 - 0.6	4.0	3.8	0.5 - 0.6	11.8	0.4
0.6 - 0.7	3.4	5.9	0.6 - 0.7	7.1	1.3
0.7 - 0.8	1.6	8.8	0.7 - 0.8	5.0	3.3
0.8 - 0.9	0.0	22.2	0.8 - 0.9	0.0	7.1
0.9 - 1.0	0.0	53.3	0.9 - 1.0	0.0	87.2

Table 3. Relative Frequency Distributions (Given in Percents) of VC and AALH for both Hyperactivated (n=322) and Non-hyperactivated (n=899) Cells

	VC			AALH	
Interval	Hyper	Non-hyper	Interval	Hyper	Non-hyper
0 - 20	0.0	0.0	0 - 2	1.6	77.5
20 - 40	0.0	7.6	2 - 4	13.3	21.4
40 - 60	4.0	25.3	4 - 6	27.3	0.9
60 - 80	6.2	21.6	6 - 8	25.5	0.1
80 - 100	11.8	16.6	8 - 10	16.8	0.0
100 - 120	19.0	10.2	10 - 12	6.8	0.1
120 - 140	18.6	9.8	12 - 14	4.7	0.0
140 - 160	14.9	5.6	14 - 16	1.5	0.0
160 - 180	8.4	3.1	16 - 18	1.9	0.0
180 - 200	4.7	0.2	18 - 20	0.3	0.0
200 -	12.4	0.0	20 -	0.3	0.0

The frequency distributions shown provide a refined description of the pattern of variability for each of the motility parameters shown. All frequency distributions were constructed using BMDP statistical software.

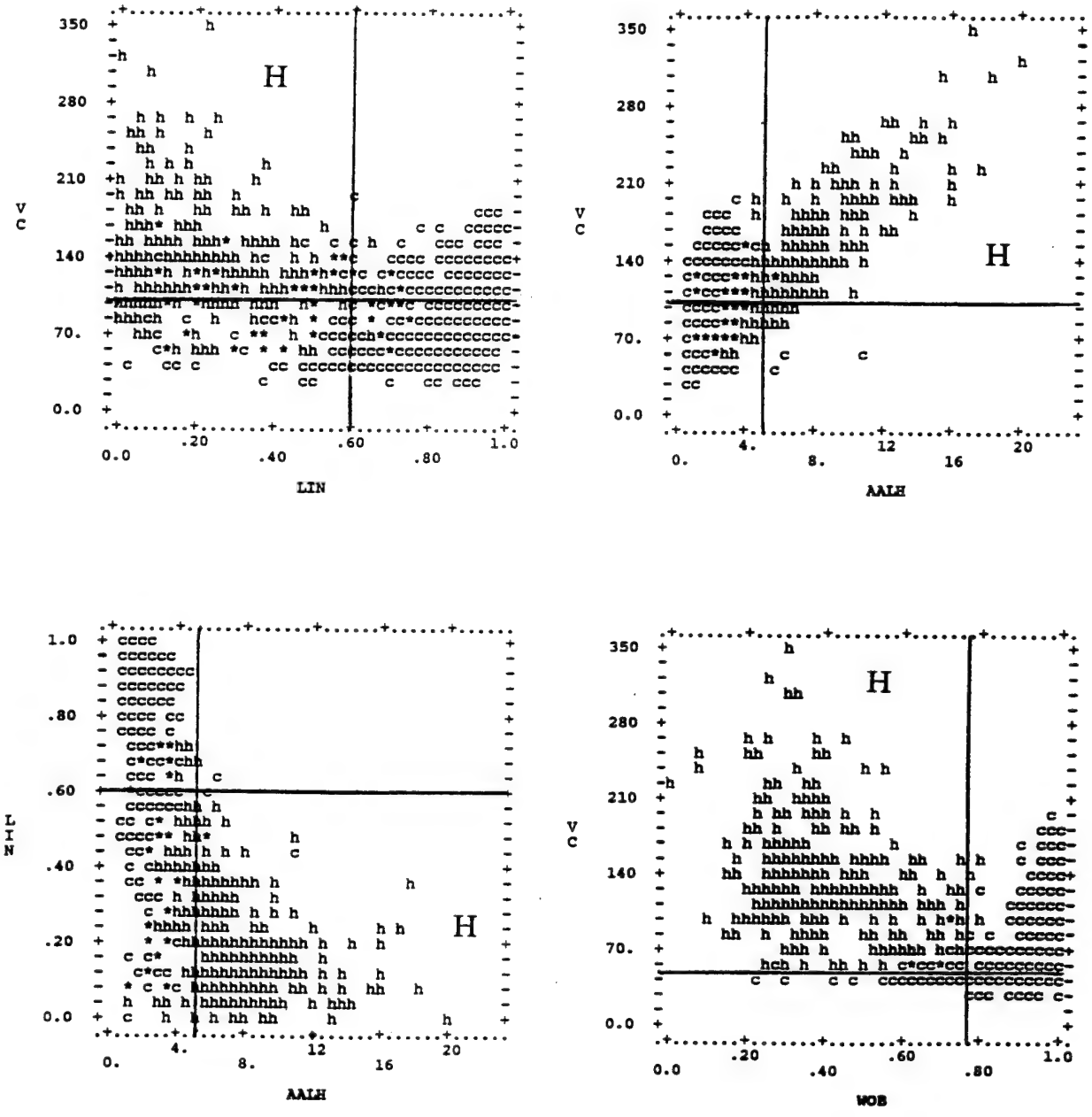


Figure 2. Scatterplots of Motion Parameters with Class Identifiers for Hyperactivated (h) and Non-hyperactivated (c). The scatterplots produced using BMDP show the degree of class separation attainable with motility parameter pairs. The symbol, c, originally represented circular or linear behavior. It was retained in this figure because it visually contrasts well with the symbol, h.

designates both classes of cells. Using the partitions suggested by Mortimer and Mortimer,[11] the first three scatterplots are divided into quadrants, with the symbol, H, denoting the quadrant where the values of both motility parameters were consistent with hyperactivated motility. For the fourth plot of VC and WOB, partitioning was achieved in a manner to be addressed later. It is apparent that in each scatterplot, the quadrant designated for hyperactivated cells contains few non-hyperactivated cells. The least pure is the partition formed on VC and LIN. For all but the VC and WOB plot, hyperactivated cells were also plentiful in other quadrants, suggesting that classification rules based on these partitions would be adequate to identify a cell as hyperactivated. However, classification rules based on these partitions would not be adequate for correctly classifying all cells in a mixed population of hyperactivated and non-hyperactivated cells. Analogous arguments for higher dimensions can be made.

Several workers have advocated the use of VC, LIN, AALH, or MALH in combination for classifying hyperactivated sperm cells.[3,4,11] These rules were based solely on the subjective extension of single motility parameters as classifiers. We have taken a comprehensive and objective analytical approach by applying regression and discriminant analysis to the problem of using multiple motility parameters to classify sperm cell motility. Discriminant analysis can be used to separate classes based on a linear compound of the motility parameters. This compound is simply a one-dimensional index that can be used to classify the observations into groups. In this simple two-group environment, discriminant analysis is analogous to performing a regression analysis on a binary (0,1) class variable and then assigning an observation to class one if the predicted value is 0.5 or greater and to class zero otherwise. The BMDP statistical software supporting both discriminant analysis and regression was used to model the relationship among motility parameters and class assignment, with more emphasis being given the regression approach.

The rationale for using both regression and discriminant analysis routines to support the derivation of motility classification rules was to offset a failure to meet the assumptions of the formal discriminant analysis and to make use of greater flexibility in the regression routines. Discriminant analysis assumes that the variables used to classify groups come from multivariate normal distributions, which differ only in location. This assumption is violated by the apparent nonnormality of many of the motility parameters (Figure 1 and appendix). The common covariance matrix assumption is also doubtful (Table 1). Without these assumptions, the computed probability of class membership for each cell is invalid. However, successful applications are possible when assumptions are violated (see Reference 12) by using the discriminant index as a measure of separation between classes, devaluing its use in forming a probability of class membership. The advantage afforded by regression is that the regression routines are more convenient for conducting variable selection and checking model adequacy. In regression, the predicted value for each cell is used as a relative score for class assignment, relying only on the assumptions usually made for a least-squares fit.

Table 4 summarizes the results of a subset of best models established using stepwise discriminant analysis and all possible subsets regression. The models are labeled discriminant 1 (D1) - discriminant 24 (D24). Three models using the motility parameters recommended[3,4,11] are given as D25-D27. Models, for a fixed number of motility characteristics, are listed in the order of decreasing R^2. The models were evaluated in terms of their efficiency, which is defined as the ability to correctly classify both hyperactivated and non-hyperactivated cells. In this analysis, classifying a hyperactivated cell as non-hyperactivated was as grievous an error as classifying a non-hyperactivated cell as hyperactivated. The percentage of correctly classified cells (hyperactivated and non-hyperactivated) was used to define efficiency. In computing this percentage resubstitution error, random cross validation and the test sample method were used. Resubstitution error was measured by establishing the classification rules based on a data set and then implementing the rule on the same data set to compute the efficiency. The obvious problem with using only resubstitution error is that there is no way to gauge the sensitivity of the established rules to variations in the data, thereby weakening claims of general applicability. Cross validation, a frequently used method to address this concern, was accomplished either by randomly targeting many subsets of the data against which to implement the rule, or by forming the rule based on a large portion (75%) of the data, and evaluating its performance against the remaining 25%. Information on the sensitivity of the classification rules to variations in the data was gathered by use of either of the cross validation approaches. During variable selection, the resubstitution error was used in determining efficiency because it allowed direct comparisons among models. Final results are reported in terms of cross validation.

Other factors important in model derivation were the needs for parsimoniousness and the avoidance of colinearity. In terms of a regression model, the explained variation or R^2 should be as high as possible consistent with the requirement that the model be simple with a minimum of measures. This is equivalent to striving for low values of Wilk's lambda in the discriminant analysis. The residuals were not to suggest a model inadequ . For example, suggesting that a quadratic expressior one of the variables would have been more appropriate. Lastly, multicolinearity, a statistical redundancy among variables in the model, was avoided to avert the danger that, although prediction may seem to improve with correlated variables in the model for the data set examined, the stability of the classifying rule for other data sets becomes suspect.

With these points in mind, Table 4 shows that models based on WOB, LIN, AALH, MALH, and VC*AALH gave efficiencies >90%. WOB was the best performer, the order being WOB > LIN > AALH > MALH > VC*AALH. The proportion of the variation associated with class distinction that is explained by WOB is 0.838. The stepwise discriminant rule established would misclassify 22 hyperactivated cells as being non-hyperactivated and 19 non-hyperactivated cells as being hyperactivated for an overall classification efficiency of 96.64%. At this stage of reporting efficiencies are given in terms of resubstitution misclassification error. Using the regression del, 31 hyperactivated cells and 15 non-hyperactivated cells were misclassified for a classification efficiency of 96.23%.

Table 4. Summary of Best Models Using Discriminant/Regression Analysis Based on 322 Hyperactivated and 899 Non-hyperactivated Cells

Model	Variables	H(missed)	NH(missed)	Efficiency (%)	R^2
D1	WOB	22 / 31	19 / 15	96.64 / 96.23	0.838
D2	LIN	21 / 34	63 / 46	93.12 / 93.45	0.702
D3	AALH	59 / 91	5 / 4	94.76 / 92.22	0.639
D4	MALH	63 / 109	16 / 9	93.53 / 90.34	0.600
D5	VC*AALH	113 / 188	4 / 0	90.42 / 84.60	0.411
D6	STR	132 / 171	78 / 36	82.80 / 83.05	0.356
D7	VC	108 / 209	205 / 56	74.37 / 78.30	0.261
D8	VST	56 / 210	301 / 26	70.76 / 80.67	0.235
D9	AALH/LIN	190 / 283	1 / 0	84.36 / 76.82	0.140
D10	WOB, AALH	21 / 32	16 / 10	96.97 / 96.56	0.856
D11	WOB, VC	23 / 30	12 / 11	97.13 / 96.64	0.847
D12	WOB, MALH	22 / 31	16 / 10	96.89 / 96.64	0.846
D13	WOB, VST	22 / 30	16 / 12	96.89 / 96.56	0.843
D14	WOB, VC*AALH	24 / 32	16 / 12	96.72 / 96.40	0.840
D15	WOB, AALH, VC*AALH	16 / 23	15 / 11	97.46 / 97.22	0.856
D16	WOB, AALH, STR	24 / 26	13 / 11	96.97 / 96.97	0.851
D17	WOB, AALH, AALH/LIN	20 / 30	16 / 10	97.05 / 96.72	0.851
D18	WOB, MALH, STR	19 / 27	15 / 13	97.22 / 96.72	0.850
D19	WOB, STR, VC	24 / 29	12 / 10	97.05 / 96.81	0.850
D20	WOB, AALH, STR, VC*AALH	15 / 22	12 / 11	97.79 / 97.30	0.860
D21	WOB, AALH, VC*AALH, VC	16 / 23	11 / 10	97.79 / 97.30	0.860
D22	WOB, LIN, STR, VC	17 / 20	14 / 11	97.46 / 97.46	0.856
D23	WOB, LIN, AALH, STR	16 / 19	14 / 13	97.54 / 97.38	0.859
D24	WOB, AALH, VC*AALH, VST	16 / 23	12 / 11	97.71 / 97.22	0.858
D25	VC, LIN, AALH	23 / 34	34 / 29	95.33 / 94.84	0.757
D26	VC, LIN, MALH	24 / 38	40 / 30	94.76 / 94.43	0.746
D27	VC, LIN, VC*AALH	24 / 40	50 / 36	93.78 / 93.78	0.729

This table shows the number of cells misclassified by each of 27 BMDP-produced models, listing the overall efficiency of classification for each.

All "best" models using two, three, or four motility parameters determined by either discriminant or regression analysis contain WOB as one of the parameters. Further, though not shown for each model, WOB was the largest contributor to explained variation for all models. The gain in efficiency by adding additional motility parameters to WOB must be considered modest. Finally, three-variable models (D25-D27) using the motility parameters most popular in the literature show an efficiency less than that achieved by WOB alone!

The large number of models shown in Table 4, which reasonably could be used to predict hyperactivated motility, was culled based on their sensitivity in predicting hyperactivated motility for other sperm samples. This involved first looking at cross validation results in misclassification. Although there was some variation in cross validation rates relative to the resubstitution rates, there were no instances so different to suggest eliminating any of the possibilities on that basis alone. The question of parsimony is largely a judgment as to how much is being contributed through adding more terms in the model and at what risk of multicolinearity. Table 5 shows the correlation structure among the motility parameters used in the models. For example, the correlation between VC*AALH and AALH is 0.933. This means that AALH is capable of explaining 87% (0.933 squared) of the variation of VC*AALH. The implication is that AALH and VC*AALH are too close statistically to be used as predictors in the same model. Similarly, a 0.904 correlation exists between WOB and LIN. They too were judged too close statistically. These results effectively eliminate the "best" four-term models (D20-D24) as well as the best of the three variable models (D15) from consideration. Considering parsimony leaves only models D10 and D11, if not just D1. Consider D10 and D11. They misclassify 37 and 35 cells, respectively. The best of the remaining three-term models misclassifies 34 cells. The slight increase in efficiency does not warrant the inclusion of a third term.

In summary, WOB, WOB and AALH, or WOB and VC are the preferred models on which to base classification rules. The motility parameters WOB and AALH are more correlated than WOB and VC, and therefore run a greater risk of inflating the standard error of prediction. Thus, the best choice would be the latter model based on WOB and VC. The regression form of that model would be Predicted Class = -0.332250 -0.000985VC + 1.456690WOB. If the predicted value for class was closer to zero than to one, codes used for hyperactivated and non-hyperactivated cells, respectively, the cell would be classified hyperactivated; otherwise, non-hyperactivated. For example, if VC=150 and WOB=0.5, then the class prediction is 0.25, indicating a hyperactivated cell. With 0.5 equidistant from the class identifiers, we may equivalently express the constraint for hyperactivity as WOB < 0.571330 + 0.000676VC. The corresponding discriminant model would indicate hyperactivity if WOB < 0.596416 + 0.000675VC. There is little difference between the approaches as long as the terms in the model have good predictive ability. Some difference would be expected, for example, with a model based on VST and AALH/LIN. The regression rule for the model WOB and AALH would be to classify a cell as hyperactivated if WOB < 0.564818 + 0.018096 AALH. A regression classification rule based on WOB alone would partition the cells at a WOB value of 0.646, with WOB being less than that value, indicating hyperactivity.

Table 5. Correlation Matrix of Motility Parameters

	VC	VST	LIN	MALH	AALH	STR	WOB	AALH/LIN	VC*AALH
VC	1.000								
VST	0.257	1.000							
LIN	-0.445	0.664	1.000						
MALH	0.701	-0.353	-0.765	1.000					
AALH	0.753	-0.338	-0.778	0.954	1.000				
STR	-0.407	0.586	0.859	-0.646	-0.645	1.000			
WOB	-0.468	0.591	0.904	-0.786	-0.812	0.685	1.000		
AALH/LIN	0.353	-0.295	-0.441	0.491	0.496	-0.527	-0.437	1.000	
VC*AALH	0.807	-0.217	-0.622	0.876	0.933	-0.548	-0.660	0.502	1.000

The correlation matrix was produced using BMDP. It helps identify those variable combinations which hold potential problems in the analysis. Strong correlations suggest a near linear dependency between variables which, if included together in a model, would act to inflate the error of prediction.

3.4 CART.

An approach distinct from the regression and discriminant analyses above is given by tree-structured classification. CART was the principal software used; FACT software was used for corroboration. Only the CART results are reported. The CART routine offers many options; only the defaults were used. Generally, for univariate splits, CART works as follows. Each possible predictor variable (motion parameter) for class is examined individually. For an individual variable, the program searches all the values, resting at each one to see how efficient it would be to partition the data into the hyperactivated and non-hyperactivated classes based on that value. (In our data set, this requires over 1200 assessments of efficiency for each variable.) The routine notes the best value for that variable based on classification efficiency. The variable, which partitions the data in the most efficient manner, is selected, and its value is used as the first partition of the data, creating two nodes, one each for hyperactivated and non-hyperactivated classification. Within each node, some cells may be misclassified. The routine then searches among the variables to further partition the two nodes to increase efficiency. Eventually, the routine settles on a decision tree for classification with maximum efficiency, subject to the constraint that tree complexity should not be great. A great advantage of tree-structured methods is that we are no longer bound by a linear model as we were in the regression and discriminant analyses, although linear combinations of variables can be considered. In running CART, all the motility parameters previously considered as possible predictors were included. The result was that CART chose only WOB and VC, with the rule: classify as hyperactivated if WOB $\leq$ 0.775 and VC $>$ 50.5. Of the 1221 cases examined, only 12 non-hyperactivated cells and 2 hyperactivated cells were misclassified for an efficiency of 98.85%. This efficiency is higher than that of any of the previously discussed models. Despite the unusually low value for VC, compared to the literature,[3,4,11] this rule has great appeal in considering the data in Figure 2. There, the incidence of non-hyperactivated cells with low WOB and low VC is high enough to cast doubt on a model based on WOB alone. In this use, VC is merely refining a classification rule based primarily on WOB.

The use of LIN, AALH, and VC was also investigated. CART did not choose to use VC. The tree was slightly more complex, having five nodes instead of three as above. The classification efficiency was 96.47%. When a model based on WOB and AALH was attempted, CART did not choose to use AALH, opting instead for a rule based only on WOB for an efficiency of 96.97%. Other runs using linear combinations of variables were attempted but resulted in more complex decision trees.

In summary, of all of the CART models examined, one of the simplest to implement was also the best. The model based on WOB and VC performed most efficiently in classifying hyperactivated and non-hyperactivated cells with the least penalty in model complexity.

3.5 Comparison of Models.

Figure 3 illustrates the decision criteria delivered by the discriminant and CART models using VC and WOB. To understand the model differences, we have partitioned the point set WOB X VC, where WOB ranges from 0.0 to 1.0, and VC ranges from 0 to 350, according to the hyperactivity decision rules for each model. A cell whose WOB and VC values locate it in a shaded region would be classified as non-hyperactivated by CART. The unshaded region corresponds to a hyperactivated classification delivered by CART. The bold line represents the discriminant model. Points falling below that line would be classified as hyperactivated; whereas, those above the line would be classified as non-hyperactivated. (The regression model is not shown but would appear nearly coincident with the discriminant model.) Within each region, we have indicated the true number of hyperactivated and non-hyperactivated cells present. From this data, one can see the similarities and differences of the model rules and assess their relative performance.

First, consider the rectangular region within which CART would classify cells as hyperactivated. Below the discriminant model there were 299 hyperactivated cells correctly classified and 3 non-hyperactivated cells incorrectly classified. In the same region but above the discriminant model, there were 21 hyperactivated cells correctly classified by CART, and 9 non-hyperactivated cells were incorrectly classified. Note that the discriminant model would have incorrectly classified the 21 hyperactivated cells while correctly classifying the 9 non-hyperactivated cells. CART is 12 cells more accurate than the discriminant model in this region. In the shaded regions (Figure 3) above the discriminant model, their performance is identical, incorrectly classifying 2 hyperactivated cells and correctly classifying 878 non-hyperactivated cells. A difference is seen again for the shaded region corresponding to low values of WOB and VC. There, the discriminant model would incorrectly classify 9 non-hyperactivated cells, bringing the CART performance advantage to 21 cells. This figure also shows that using VC to establish a lower threshold is beneficial in improving a classification by WOB alone. Twenty-three cells would have been incorrectly classified using a WOB criterion without considering VC. Our inference is that WOB is a stable measure and good classifier except for very slow moving cells, which WOB sometimes errantly classifies as hyperactivated.

3.6 Sensitivity Analysis.

Earlier, we stressed the undesirability of forming a model based on a data set and evaluating the efficiency of the model based on that same data set. Thus far, to facilitate model comparison, the resubstitution error has been used for computing efficiency. However, it should be noted that several different methods of cross validation, including jackknifed estimates, random subsets, and the test sample method were also employed. In general, we found that the best linear models and CART were resistant to changes in the data from cross validation efforts. The efficiency according to cross validation among the various methods was, at worst, 98%.

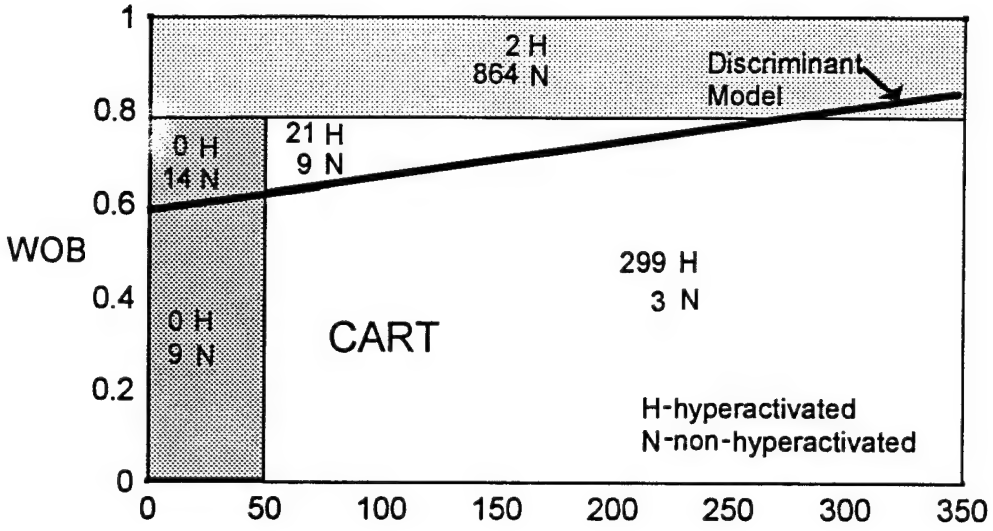


Figure 3. A comparison of discriminant and CART models to classify hyperactivation. The figure shows the model rules for classifying cells and indicates the correctness of those classifications. The white region is the hyperactivated region for CART. The hyperactivated region for the discriminant model is the area below the bold line.

3.7 Motion Analysis System (CellTrak).

The previous analysis was based on the motion parameter values as measured by the CellSoft system. To determine how the resulting model performs when applied to parameter values measured by another system, the tapes were reanalyzed using the Motion Analysis System. Although the tapes were the same, it was impossible to determine if exactly the same cells were being analyzed. The number of cells analyzed was 1119, somewhat less than the 1221 analyzed by CellSoft. Table 6 summarizes the individual parameters for the Motion Analysis System. This investigation shows that some motility parameters are different than those reported by CellSoft. The largest difference occurs with AALH. CellTrak seems to approximately double the AALH values relative to CellSoft. Another difference noted is with VST values, particularly among the non-hyperactivated cells. The VST as measured by CellTrak is approximately 12 μm/s slower for non-hyperactivated cells as that measured by CellSoft. Other differences include means for LIN, and WOB for the non-hyperactivated cells and predictably means for AALH/LIN and VC*AALH for all cells. Still, scatter plot examination (not shown) reveals a similar data structure between parameters to that observed with CellSoft. Implementation of the CART decision rule based on the CellSoft data to the data produced by CellTrak yielded surprisingly good results. Forty cells were misclassified for an efficiency of 96.4%. In an effort to calibrate the model for the system being used, CART was performed on the CellTrak data to determine a model best suited for classifying this new data. CART again picked WOB and VC together with the same tree structure to predict hyperactivity! The rule, only slightly different than that for CellSoft, would be to classify as hyperactive cells showing WOB $\leq$ 0.705 and VC $>$ 49.2. The number of cells misclassified was 30 for an efficiency of 97.3%. Cross validation results reported efficiencies, at worst, of 97%.

3.8 Data Cleansing.

A further check on the model validity involved reexamining each of the cell tracks analyzed by CellSoft and CellTrak. Sperm cells that were in the gray area for hyperactivated motility were removed from the data sets, and the CART routine was repeated for both the CellSoft and CellTrak results. With CellSoft, 13 cells were removed, and no change at all was recorded for the decision rule values of 0.775 for WOB and 50.5 μm/s for VC. For CellTrak, 40 cells were removed with a slight change in values. The WOB partition changed from 0.705 to 0.685, and the VC partition changed from 49.2 μm/s to 54.9 μm/s. The new efficiencies for CellSoft and CellTrak were 98.7% and 98.4%, respectively, computed as a cross validation efficiency.

4. CONCLUSIONS

Overall, the CART model based on VC and WOB is preferred. It certainly performs better than the discriminant or regression models based on the same motility

Table 6. Summary Statistics for the Motility Parameters Determined by CellTrak for both Hyperactivated and Non-hyperactivated Cells

Motility Parameter	Class	Mean ± SD	Sample Minimum	Maximum	Range
VC	hyper	136.9 ± 51.1	50.4	383.5	333.2
	non-hyper	77.9 ± 37.0	25.0	186.0	161.0
VST	hyper	28.5 ± 18.3	0.0	109.3	109.3
	non-hyper	57.9 ± 38.8	0.0	174.0	174.0
LIN	hyper	0.24 ± 0.17	0.01	0.82	0.81
	non-hyper	0.73 ± 0.26	0.01	0.99	0.98
AALH	hyper	14.3 ± 6.5	1.5	63.0	61.5
	non-hyper	3.9 ± 2.1	1.3	13.0	11.7
STR	hyper	0.53 ± 0.27	0.0	0.99	0.99
	non-hyper	0.84 ± 0.22	0.0	0.99	0.99
WOB	hyper	0.42 ± 0.15	0.14	0.90	0.76
	non-hyper	0.84 ± 0.21	0.12	0.99	0.77
AALH/LIN	hyper	122.1 ± 156.7	4.4	1575.8	1571.4
	non-hyper	11.3 ± 31.9	1.4	390.0	388.6
VC*AALH	hyper	2153.2 ± 1708.3	223.4	12434.4	12211.0
	non-hyper	327.5 ± 290.1	38.0	2275.0	2237.0

Summary statistics for each motility parameter are reported individually for each cell state. The mean ± the sample standard deviation, for the population, gives information as to the location of the majority of motility parameter values. The minimum, maximum, and range provide information as to the extremes. All summary statistics were computed using BMDP statistical software.

parameters, and much better than linear, discriminant, linear regression, or CART models based on the motility parameters most commonly used in the literature.[3,4,11]

The classification rule for analysis with the CellSoft system is WOB $\leq$ 0.775 and VC $\geq$ 51 μm/s. For the CellTrak system, the rule is WOB $\leq$ 0.705 and VC $\geq$ 50 μm/s, or for a more restricted classification WOB $\leq$ 0.685 and VC $\geq$ 55 μm/s.

LITERATURE CITED

1. Young, R.J., and Heitkamp, D.H., Inhibition of Rabbit Sperm Cell Hyperactivated Motility by Metallic Ions in Toxicological Testing, ERDEC-TR-226, U.S. Army Edgewood Research, Development and Engineering Center, Aberdeen Proving Ground, MD, March 1995, UNCLASSIFIED Report.

2. Murad, C., DE Lamirande, E., and Gagnon, C., "Hyperactivated Motility is Coupled with Interdependent Modifications at Axonemal and Cytosolic Levels in Human Spermatozoa," J. Androl. Vol. 13, pp 323-331 (1992).

3. Burkman, L.J., "Discrimination Between Nonhyperactivated and Classical Hyperactivated Motility Patterns in Human Spermatozoa Using Computerized Analysis," Fertil. Steril. Vol. 55, pp 363-371 (1992).

4. Robertson, L., Wolf, D.P., and Tash, J.S., "Temporal Changes in Motility Parameters Related to Acrosomal Status: Identification and Characterization of Populations of Hyperactivated Human Sperm," Biol. Reprod. Vol. 39, pp 797-805 (1988).

5. Young, R.J., Bodt, B.A., Iturralde, T.G., and Starke, W.C., "Automated Analysis of Rabbit Sperm Motility and the Effect of Chemicals on Sperm Motion Parameters," Mol. Reprod. Dev. Vol. 33, pp 347-356 (1992).

6. Young, R.J., New Medium for the Culture of Rabbit Sperm for Toxicology Testing, ERDEC-TR-139, U.S. Army Edgewood Research, Development and Engineering Center, Aberdeen Proving Ground, MD, December 1993, UNCLASSIFIED Report (AD A277 504).

7. Young, R.J., Starke, W.C., Bodt, B.A., and Laurie, E.A., Statistical Validation of the CellSoft Motion Analysis System for the Study of the Motility Characteristics of Rabbit Sperm Cells, CRDEC-TR-214, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, January 1991, UNCLASSIFIED Report (AD B153 040L).

8. Yanagimachi, R., "Mammalian Fertilization," In The Physiology of Reproduction, pp 135-185, E. Knobil and J. Neill, Eds., Raven Press, 1988.

9. Yanagimachi, R., "The Movement of Golden Hamster Spermatozoa Before and After Capacitation," J. Reprod. Fertil. Vol. 23, pp 193-196 (1970).

10. Conover, W.J., Practical Nonparametric Statistics, Chapter 5, 2nd ed., John Wiley and Sons, New York, NY, 1980.

11. Mortimer, S.T., and Mortimer, D., "Kinematics of Human Spermatozoa Incubated Under Capacitating Conditions," J. Androl. Vol. 11, pp 195-203 (1990).

12. Breiman, L., Friedman, J., Olshen, R., and Stone, C., Classification and Regression Trees, Wadsworth, Incorporated, Belmont, CA, 1984.

APPENDIX

MOTILITY PARAMETER DISTRIBUTIONS (BY HYPERACTIVITY)

HISTOGRAM OF VARIABLE 3 VC

	SYMBOL	COUNT	MEAN	ST.DEV.
hyper	A	322	137.566	51.966
nonhyper	B	899	83.052	35.710

EACH SYMBOL REPRESENTS 5 OBSERVATIONS

INTERVAL NAME	25 50 75 100 125 150 175 200	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
*26.25	+	2	2	0.2	0.2
*35	+BB	9	11	0.7	0.9
*43.75	+BBBBBBBBBBBBBBBBBBB	99	110	8.1	9.0
*52.5	+BBBBBBBBBBBBBBBBBBB	96	206	7.9	16.9
*61.25	+AABBBBBBBBBBBBBBBBBBB	111	317	9.1	26.0
*70	+AABBBBBBBBBBBBBBBBBBB	110	427	9.0	35.0
*78.75	+AABBBBBBBBBBBBBBBB	89	516	7.3	42.3
*87.5	+AABBBBBBBBBBBBBB	78	594	6.4	48.6
*96.25	+AAAABBBBBBBBBBBBB	86	680	7.0	55.7
*105	+AAAABBBBBBBBBB	72	752	5.9	61.6
*113.75	+AAAAAAABBBBBBB	71	823	5.8	67.4
*122.5	+AAAAABBBBBBB	61	884	5.0	72.4
*131.25	+AAAAAABBBBBBB	66	950	5.4	77.8
*140	+AAAAABBBBBBB	61	1011	5.0	82.8
*148.75	+AAAABBBBBB	50	1061	4.1	86.9
*157.5	+AAAABBB	33	1094	2.7	89.6
*166.25	+AAABBBB	37	1131	3.0	92.6
*175	+AABB	20	1151	1.6	94.3
*183.75	+AAB	16	1167	1.3	95.6
*192.5	+B	3	1170	0.2	95.8
*201.25	+AAA	13	1183	1.1	96.9
*210	+AA	11	1194	0.9	97.8
*218.75	+A	4	1198	0.3	98.1
*227.5	+A	4	1202	0.3	98.4
*236.25	+	2	1204	0.2	98.6
*245	+A	3	1207	0.2	98.9
*253.75	+A	4	1211	0.3	99.2
*262.5	+	2	1213	0.2	99.3
*271.25	+	2	1215	0.2	99.5
*280	+	1	1216	0.1	99.6
*288.75	+	0	1216	0.0	99.6
*297.5	+	0	1216	0.0	99.6
*306.25	+	0	1216	0.0	99.6
*315	+	2	1218	0.2	99.8
*323.75	+	1	1219	0.1	99.8
*332.5	+	0	1219	0.0	99.8
*341.25	+	0	1219	0.0	99.8
*350	+	2	1221	0.2	100.0
	25 50 75 100 125 150 175 200				

HISTOGRAM OF VARIABLE 4 VST

	SYMBOL	COUNT	MEAN	ST.DEV.
hyper	A	322	30.444	21.120
nonhyper	B	897	71.042	35.494

EACH SYMBOL REPRESENTS 5 OBSERVATIONS

INTERVAL NAME		FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
*0	+	0	0	0.0	0.0
*5	+AAAAAA	30	30	2.5	2.5
*10	+AAAAAAB	37	67	3.0	5.5
*15	+AAAAAAAB	41	108	3.4	8.9
*20	+AAAABB	30	138	2.5	11.3
*25	+AAAAAAABBBBB	59	197	4.8	16.2
*30	+AAAAAAABBBB	53	250	4.3	20.5
*35	+AAAABBBBBBBBBBB	77	327	6.3	26.8
*40	+AAAABBBBBBBBBBB	76	403	6.2	33.1
*45	+AAAABBBBBBBBBBB	77	480	6.3	39.4
*50	+AAABBBBBBBBBBBBB	80	560	6.6	45.9
*55	+AABBBBBBBBBBB	66	626	5.4	51.4
*60	+AAABBBBBBBBBBB	69	695	5.7	57.0
*65	+ABBBBBBBBBB	55	750	4.5	61.5
*70	+AABBBBBBBBBBB	67	817	5.5	67.0
*75	+ABBBBBBBBB	50	867	4.1	71.1
*80	+ABBBBBBBB	43	910	3.5	74.7
*85	+BBBBBBB	34	944	2.8	77.4
*90	+ABBBBBB	34	978	2.8	80.2
*95	+BBBBBB	32	1010	2.6	82.9
*100	+BBBBBB	31	1041	2.5	85.4
*105	+BBB	17	1058	1.4	86.8
*110	+BBB	17	1075	1.4	88.2
*115	+BBBB	19	1094	1.6	89.7
*120	+BBB	17	1111	1.4	91.1
*125	+BBBB	18	1129	1.5	92.6
*130	+BBB	15	1144	1.2	93.8
*135	+BBB	16	1160	1.3	95.2
*140	+BBB	13	1173	1.1	96.2
*145	+BB	9	1182	0.7	97.0
*150	+BB	8	1190	0.7	97.6
*155	+B	7	1197	0.6	98.2
*160	+BB	9	1206	0.7	98.9
*165	+BB	8	1214	0.7	99.6
*170	+	2	1216	0.2	99.8
*175	+	2	1218	0.2	99.9
*180	+	1	1219	0.1	100.0
*185	+	0	1219	0.0	100.0

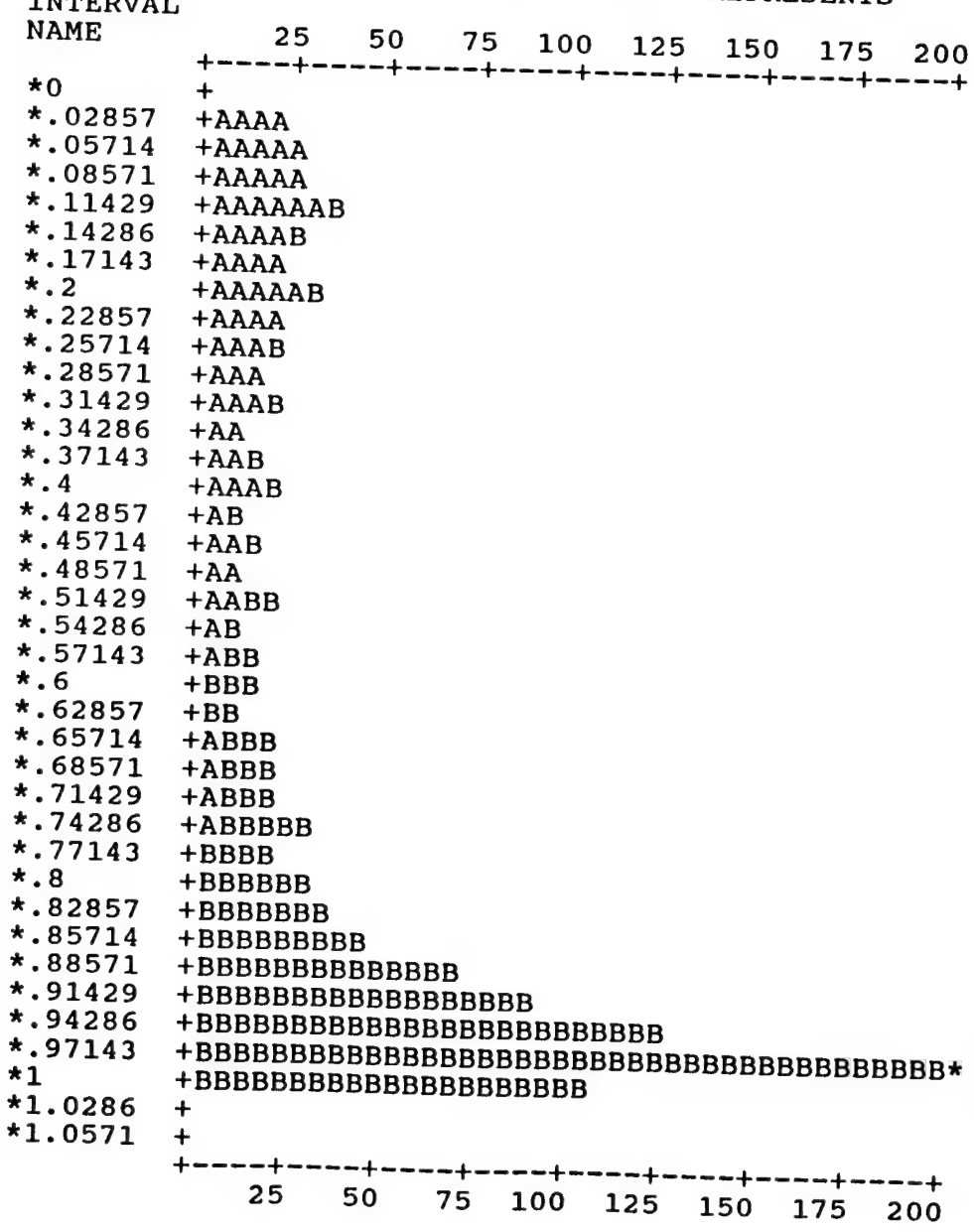
25 50 75 100 125 150 175 200

HISTOGRAM OF VARIABLE 5 LIN

	SYMBOL	COUNT
hyper	A	322
nonhyper	B	899

EACH SYMBOL REPRESENTS 5 OBSERVATIONS

MEAN	ST.DEV.
0.241	0.176
0.845	0.177



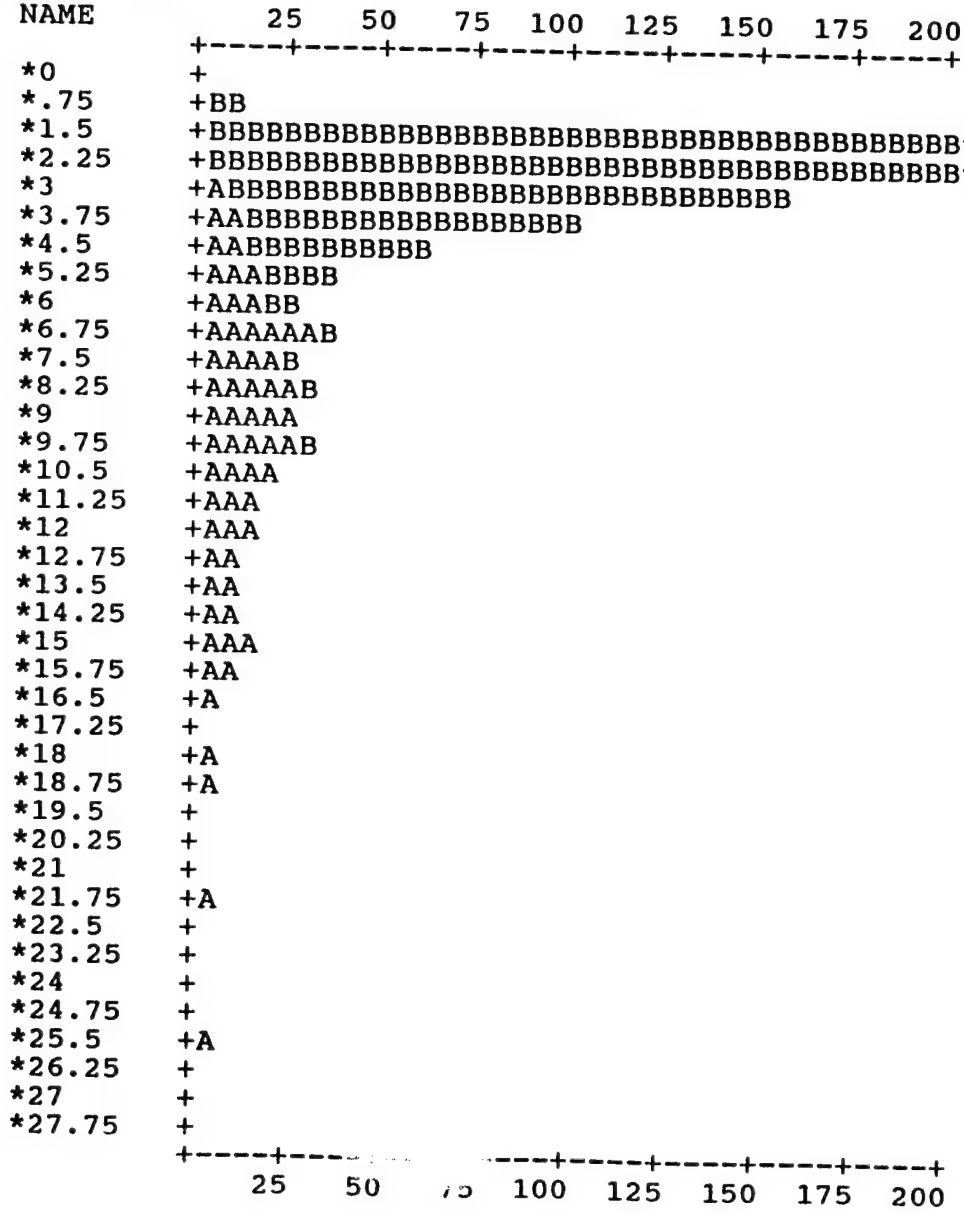
INTERVAL NAME	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
*0	0	0	0.0	0.0
*.02857	19	19	1.6	1.6
*.05714	23	42	1.9	3.4
*.08571	26	68	2.1	5.6
*.11429	35	103	2.9	8.4
*.14286	26	129	2.1	10.6
*.17143	20	149	1.6	12.2
*.2	29	178	2.4	14.6
*.22857	20	198	1.6	16.2
*.25714	19	217	1.6	17.8
*.28571	16	233	1.3	19.1
*.31429	18	251	1.5	20.6
*.34286	12	263	1.0	21.5
*.37143	16	279	1.3	22.9
*.4	18	297	1.5	24.3
*.42857	10	307	0.8	25.1
*.45714	15	322	1.2	26.4
*.48571	12	334	1.0	27.4
*.51429	20	354	1.6	29.0
*.54286	10	364	0.8	29.8
*.57143	14	378	1.1	31.0
*.6	16	394	1.3	32.3
*.62857	9	403	0.7	33.0
*.65714	19	422	1.6	34.6
*.68571	20	442	1.6	36.2
*.71429	22	464	1.8	38.0
*.74286	30	494	2.5	40.5
*.77143	19	513	1.6	42.0
*.8	29	542	2.4	44.4
*.82857	33	575	2.7	47.1
*.85714	47	622	3.8	50.9
*.88571	70	692	5.7	56.7
*.91429	88	780	7.2	63.9
*.94286	127	907	10.4	74.3
*.97143	208	1115	17.0	91.3
*1	106	1221	8.7	100.0
*1.0286	0	1221	0.0	100.0
*1.0571	0	1221	0.0	100.0

HISTOGRAM OF VARIABLE 6 MALH

	SYMBOL	COUNT
hyper	A	321
nonhyper	B	897

EACH SYMBOL REPRESENTS 5 OBSERVATIONS

MEAN	ST.DEV.
9.858	4.834
2.256	1.279



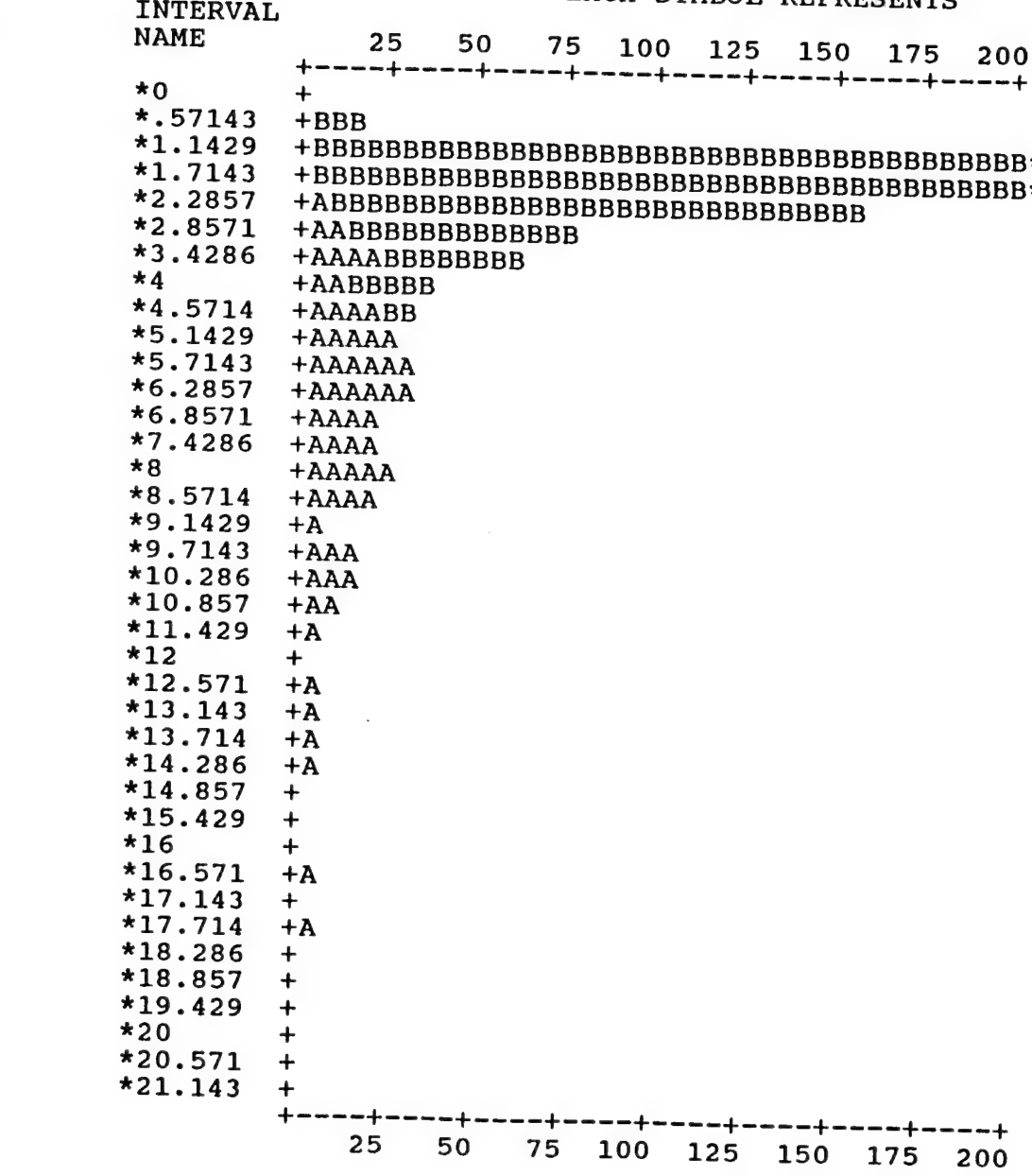
INTERVAL NAME	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
*0	0	0	0.0	0.0
*.75	9	9	0.7	0.7
*1.5	271	280	22.2	23.0
*2.25	285	565	23.4	46.4
*3	157	722	12.9	59.3
*3.75	98	820	8.0	67.3
*4.5	58	878	4.8	72.1
*5.25	37	915	3.0	75.1
*6	27	942	2.2	77.3
*6.75	35	977	2.9	80.2
*7.5	24	1001	2.0	82.2
*8.25	28	1029	2.3	84.5
*9	25	1054	2.1	86.5
*9.75	28	1082	2.3	88.8
*10.5	21	1103	1.7	90.6
*11.25	16	1119	1.3	91.9
*12	13	1132	1.1	92.9
*12.75	11	1143	0.9	93.8
*13.5	11	1154	0.9	94.7
*14.25	8	1162	0.7	95.4
*15	15	1177	1.2	96.6
*15.75	11	1188	0.9	97.5
*16.5	3	1191	0.2	97.8
*17.25	2	1193	0.2	97.9
*18	4	1197	0.3	98.3
*18.75	4	1201	0.3	98.6
*19.5	1	1202	0.1	98.7
*20.25	1	1203	0.1	98.8
*21	1	1204	0.1	98.9
*21.75	4	1208	0.3	99.2
*22.5	1	1209	0.1	99.3
*23.25	2	1211	0.2	99.4
*24	0	1211	0.0	99.4
*24.75	1	1212	0.1	99.5
*25.5	3	1215	0.2	99.8
*26.25	2	1217	0.2	99.9
*27	1	1218	0.1	100.0
*27.75	0	1218	0.0	100.0

HISTOGRAM OF VARIABLE 7 AALH

	SYMBOL	COUNT
hyper	A	322
nonhyper	B	899

EACH SYMBOL REPRESENTS 5 OBSERVATIONS

MEAN	ST.DEV.
7.117	3.292
1.554	0.858



FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
0	0	0.0	0.0
15	15	1.2	1.2
324	339	26.5	27.8
274	613	22.4	50.2
153	766	12.5	62.7
75	841	6.1	68.9
61	902	5.0	73.9
35	937	2.9	76.7
28	965	2.3	79.0
24	989	2.0	81.0
29	1018	2.4	83.4
31	1049	2.5	85.9
22	1071	1.8	87.7
22	1093	1.8	89.5
23	1116	1.9	91.4
20	1136	1.6	93.0
7	1143	0.6	93.6
16	1159	1.3	94.9
14	1173	1.1	96.1
12	1185	1.0	97.1
6	1191	0.5	97.5
2	1193	0.2	97.7
4	1197	0.3	98.0
3	1200	0.2	98.3
4	1204	0.3	98.6
5	1209	0.4	99.0
1	1210	0.1	99.1
2	1212	0.2	99.3
1	1213	0.1	99.3
3	1216	0.2	99.6
0	1216	0.0	99.6
3	1219	0.2	99.8
1	1220	0.1	99.9
0	1220	0.0	99.9
0	1220	0.0	99.9
0	1220	0.0	99.9
1	1221	0.1	100.0
0	1221	0.0	100.0

HISTOGRAM OF VARIABLE 8 BCF

	SYMBOL	COUNT	MEAN	ST.DEV.
hyper	A	321	12.201	5.545
nonhyper	B	898	15.086	3.800

EACH SYMBOL REPRESENTS 5 OBSERVATIONS

INTERVAL NAME	25 50 75 100 125 150 175 200	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
*1.4286	+	2	2	0.2	0.2
*2.1429	+A	4	6	0.3	0.5
*2.8571	+A	5	11	0.4	0.9
*3.5714	+A	4	15	0.3	1.2
*4.2857	+AAA	14	29	1.1	2.4
*5	+AAA	16	45	1.3	3.7
*5.7143	+A	5	50	0.4	4.1
*6.4286	+AABB	19	69	1.6	5.7
*7.1429	+A	5	74	0.4	6.1
*7.8571	+AAABB	24	98	2.0	8.0
*8.5714	+AAAABBBBBBBB	59	157	4.8	12.9
*9.2857	+AAAABB	28	185	2.3	15.2
*10	+AAAABBBB	42	227	3.4	18.6
*10.714	+BBB	14	241	1.1	19.8
*11.429	+AAABBBBBBB	49	290	4.0	23.8
*12.143	+AAAAABBBBBBBBBBB	85	375	7.0	30.8
*12.857	+BBB	13	388	1.1	31.8
*13.571	+AAAABBBBBBBBBBBBBB	96	484	7.9	39.7
*14.286	+ABBB	20	504	1.6	41.3
*15	+AAAABBBBBBBBBBBBBBBBBBBBB	131	635	10.7	52.1
*15.714	+AAABBBBBBBBBBBBBBBBBB	113	748	9.3	61.4
*16.429	+ABBBBB	30	778	2.5	63.8
*17.143	+AAABBBBBBBBBBBBBBBBBBBBB	123	901	10.1	73.9
*17.857	+AABBBBB	34	935	2.8	76.7
*18.571	+ABBBBBBBBBBBBBB	79	1014	6.5	83.2
*19.286	+AABBBBBBBBB	61	1075	5.0	88.2
*20	+ABB	17	1092	1.4	89.6
*20.714	+AAABBBBBBBBB	64	1156	5.3	94.8
*21.429	+B	4	1160	0.3	95.2
*22.143	+ABBBB	27	1187	2.2	97.4
*22.857	+AB	12	1199	1.0	98.4
*23.571	+A	5	1204	0.4	98.8
*24.286	+B	7	1211	0.6	99.3
*25	+	2	1213	0.2	99.5
*25.714	+B	4	1217	0.3	99.8
*26.429	+	1	1218	0.1	99.9
*27.143	+	0	1218	0.0	99.9
*27.857	+	1	1219	0.1	100.0
	25 50 75 100 125 150 175 200				

HISTOGRAM OF VARIABLE 9 STR

	SYMBOL	COUNT
hyper	A	322
nonhyper	B	899

EACH SYMBOL REPRESENTS 5 OBSERVATIONS

MEAN	ST.DEV.
0.576	0.297
0.901	0.137

INTERVAL NAME	25 50 75 100 125 150 175 200	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
*0	+				
*.02857	+	0	0	0.0	0.0
*.05714	+A	2	2	0.2	0.2
*.08571	+AB	5	7	0.4	0.6
*.11429	+AAA	9	16	0.7	1.3
*.14286	+AA	13	29	1.1	2.4
*.17143	+A	9	38	0.7	3.1
*.2	+AAA	7	45	0.6	3.7
*.22857	+AB	17	62	1.4	5.1
*.25714	+A	8	70	0.7	5.7
*.28571	+AA	7	77	0.6	6.3
*.31429	+AA	11	88	0.9	7.2
*.34286	+A	11	99	0.9	8.1
*.37143	+A	7	106	0.6	8.7
*.4	+AAB	3	109	0.2	8.9
*.42857	+A	15	124	1.2	10.2
*.45714	+AB	5	129	0.4	10.6
*.48571	+A	11	140	0.9	11.5
*.51429	+AAB	5	145	0.4	11.9
*.54286	+AB	14	159	1.1	13.0
*.57143	+AA	10	169	0.8	13.8
*.6	+AA	10	179	0.8	14.7
*.62857	+B	11	190	0.9	15.6
*.65714	+AAABB	6	196	0.5	16.1
*.68571	+AABB	25	221	2.0	18.1
*.71429	+AABB	20	241	1.6	19.7
*.74286	+ABBB	19	260	1.6	21.3
*.77143	+AABBBBB	18	278	1.5	22.8
*.8	+AAABBBBBB	33	311	2.7	25.5
*.82857	+AABB	44	355	3.6	29.1
*.85714	+AAAABBBBB	21	376	1.7	30.8
*.88571	+AAABBBBBBBBBBB	46	422	3.8	34.6
*.91429	+AAABBBBBBBBBBBBBBBB	72	494	5.9	40.5
*.94286	+AABBBBBBBBBBBBBBBBBBB	98	592	8.0	48.5
*.97143	+AAAABBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB*	114	706	9.3	57.8
*1	+AABBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB*	229	935	18.8	76.6
*1.0286	+	286	1221	23.4	100.0
*1.0571	+	0	1221	0.0	100.0
		0	1221	0.0	100.0

HISTOGRAM OF VARIABLE 10 wob

	SYMBOL	COUNT	MEAN	ST.DEV.
hyper	A	322	0.399	0.154
nonhyper	B	899	0.943	0.082

EACH SYMBOL REPRESENTS 5 OBSERVATIONS

INTERVAL NAME	25 50 75 100 125 150 175 200	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
*0	+				
*.02857	+	0	0	0.0	0.0
*.05714	+	1	1	0.1	0.1
*.08571	+	0	1	0.0	0.1
*.11429	+	2	3	0.2	0.2
*.14286	+	1	4	0.1	0.3
*.17143	+A	2	6	0.2	0.5
*.2	+AAA	4	10	0.3	0.8
*.22857	+AA	13	23	1.1	1.9
*.25714	+AAAAA	10	33	0.8	2.7
*.28571	+AAAAA	24	57	2.0	4.7
*.31429	+AAAAAA	26	83	2.1	6.8
*.34286	+AAAAAA	30	113	2.5	9.3
*.37143	+AAAAA	28	141	2.3	11.5
*.4	+AAAAA	24	165	2.0	13.5
*.42857	+AAA	25	190	2.0	15.6
*.45714	+AAAAA	14	204	1.1	16.7
*.48571	+AAB	26	230	2.1	18.8
*.51429	+AAA	13	243	1.1	19.9
*.54286	+AAA	16	259	1.3	21.2
*.57143	+AA	15	274	1.2	22.4
*.6	+AB	10	284	0.8	23.3
*.62857	+A	9	293	0.7	24.0
*.65714	+AA	7	300	0.6	24.6
*.68571	+AB	11	311	0.9	25.5
*.71429	+AB	11	322	0.9	26.4
*.74286	+ABB	9	331	0.7	27.1
*.77143	+AB	14	345	1.1	28.3
*.8	+BBBB	10	355	0.8	29.1
*.82857	+B	18	373	1.5	30.5
*.85714	+BBB	4	377	0.3	30.9
*.88571	+BBBB	16	393	1.3	32.2
*.91429	+BBBBBBBBB	21	414	1.7	33.9
*.94286	+BBBBBBBBBBBBBBBBBBBBBBBBB	43	457	3.5	37.4
*.97143	+BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB	128	585	10.5	47.9
*1	+BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB*	300	885	24.6	72.5
*1.0286	+	336	1221	27.5	100.0
*1.0571	+	0	1221	0.0	100.0
		0	1221	0.0	100.0
	25 50 75 100 125 150 175 200				

```
HISTOGRAM OF VARIABLE   13 AALH/LIN
                              SYMBOL  COUNT         MEAN      ST.DEV.
                     hyper      A      322          92.464     192.173
                     nonhyper   B      899           2.392       3.707
INTERVAL             EACH SYMBOL REPRESENTS            5 OBSERVATIONS
NAME          25   50   75  100  125  150  175  200  FREQUENCY PERCENTAGE
         +----+----+----+----+----+----+----+----+  INT. CUM.  INT.  CUM.
*0       +
*57.143  +AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA*1107 1107  90.7  90.7
*114.29  +AAAAAAAAAAAA                               61 1168   5.0  95.7
*171.43  +AAA                                        13 1181   1.1  96.7
*228.57  +AA                                         10 1191   0.8  97.5
*285.71  +AA                                          9 1200   0.7  98.3
*342.86  +A                                           4 1204   0.3  98.6
*400     +A                                           3 1207   0.2  98.9
*457.14  +                                            0 1207   0.0  98.9
*514.29  +                                            2 1209   0.2  99.0
*571.43  +A                                           3 1212   0.2  99.3
*628.57  +                                            0 1212   0.0  99.3
*685.71  +                                            1 1213   0.1  99.3
*742.86  +                                            0 1213   0.0  99.3
*800     +A                                           3 1216   0.2  99.6
*857.14  +                                            0 1216   0.0  99.6
*914.29  +                                            0 1216   0.0  99.6
*971.43  +                                            2 1218   0.2  99.8
*1028.6  +                                            1 1219   0.1  99.8
*1085.7  +                                            0 1219   0.0  99.8
*1142.9  +                                            0 1219   0.0  99.8
*1200    +                                            0 1219   0.0  99.8
*1257.1  +                                            0 1219   0.0  99.8
*1314.3  +                                            0 1219   0.0  99.8
*1371.4  +                                            1 1220   0.1  99.9
*1428.6  +                                            0 1220   0.0  99.9
*1485.7  +                                            0 1220   0.0  99.9
*1542.9  +                                            0 1220   0.0  99.9
*1600    +                                            0 1220   0.0  99.9
*1657.1  +                                            0 1220   0.0  99.9
*1714.3  +                                            0 1220   0.0  99.9
*1771.4  +                                            0 1220   0.0  99.9
*1828.6  +                                            0 1220   0.0  99.9
*1885.7  +                                            0 1220   0.0  99.9
*1942.9  +                                            0 1220   0.0  99.9
*2000    +                                            0 1220   0.0  99.9
*2057.1  +                                            1 1221   0.1 100.0
*2114.3  +                                            0 1221   0.0 100.0
         +----+----+----+----+----+----+----+----+
              25   50   75  100  125  150  175  200
```

HISTOGRAM OF VARIABLE 14 VC*AALH

	SYMBOL	COUNT	MEAN	ST.DEV.
hyper	A	322	1124.312	984.320
nonhyper	B	899	144.853	128.281

EACH SYMBOL REPRESENTS 5 OBSERVATIONS

INTERVAL NAME		FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
*0	+	0	0	0.0	0.0
*180	+AAABBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB*	681	681	55.8	55.8
*360	+AAAAAAABBBBBBBBBBBBBBBBBBBBBBBBBBBBBB	193	874	15.8	71.6
*540	+AAAAAAAABBBBBBBBBBBB	100	974	8.2	79.8
*720	+AAAAAAAAABBB	62	1036	5.1	84.8
*900	+AAAAAAAA	40	1076	3.3	88.1
*1080	+AAAAA	24	1100	2.0	90.1
*1260	+AAAAA	25	1125	2.0	92.1
*1440	+AAAA	21	1146	1.7	93.9
*1620	+AAA	14	1160	1.1	95.0
*1800	+AA	8	1168	0.7	95.7
*1980	+AA	9	1177	0.7	96.4
*2160	+AA	9	1186	0.7	97.1
*2340	+A	3	1189	0.2	97.4
*2520	+A	3	1192	0.2	97.6
*2700	+A	6	1198	0.5	98.1
*2880	+A	5	1203	0.4	98.5
*3060	+	1	1204	0.1	98.6
*3240	+	2	1206	0.2	98.8
*3420	+A	4	1210	0.3	99.1
*3600	+	0	1210	0.0	99.1
*3780	+	2	1212	0.2	99.3
*3960	+A	3	1215	0.2	99.5
*4140	+	0	1215	0.0	99.5
*4320	+	0	1215	0.0	99.5
*4500	+	1	1216	0.1	99.6
*4680	+	0	1216	0.0	99.6
*4860	+	1	1217	0.1	99.7
*5040	+	0	1217	0.0	99.7
*5220	+	0	1217	0.0	99.7
*5400	+	0	1217	0.0	99.7
*5580	+	0	1217	0.0	99.7
*5760	+	1	1218	0.1	99.8
*5940	+	0	1218	0.0	99.8
*6120	+	2	1220	0.2	99.9
*6300	+	0	1220	0.0	99.9
*6480	+	1	1221	0.1	100.0
*6660	+	0	1221	0.0	100.0

25 50 75 100 125 150 175 200